OPEN ACCESS

2nd International Electronic Conference on Sensors and Applications
15–30 November 2015

www.mdpi.com/journal/sensors

Conference Proceedings Paper – Sensors and Applications

Application of J48 Decision Tree for the Identification of Water bodies Using Landsat 8 OLI Imagery

Tri Dev Acharya, In Tae Yang and Dong Ha Lee *

Department of Civil Engineering, Kangwon National University, Chuncheon 200-701, Korea;
E-Mails: tridevacharya@gmail.com (T.D.A.); intae@kangwon.ac.kr (I.T.Y)

* Author to whom correspondence should be addressed; E-Mail: geodesy@kangwon.ac.kr;
Tel.: +82-33-250-6235; Fax: +82-33-259-5543.

Published: 10 November 2015

Abstract: Water bodies are essential to humans and other forms of life. Identification of water bodies can be useful in various ways: estimation of water availability, demarcation of flooded regions and so on. In past decades, Landsat sensors have been used for land use classification using various unsupervised and supervise methods. With the introduction of new OLI sensor in Landsat 8 with improved qualities, the accuracy of classification has been much improved. With increasing quality, the data size are also increasing, at the same time data mining techniques are developed to improve the classification efficiencies. The objective of the study is to apply J48 decision tree to identify water bodies using Landsat 8 OLI imageries. J48 is an open source java implementation of C4.5 decision tree. The imagery for the study is from Gangwon-do area, Republic of Korea. Training data with individual bands and band ratios were used to develop the decision tree model and later applied to the whole study area. The performance of the result was statically analyzed using *Kappa statistics* and *Area under Curve*. The result shows a successful application of data mining technique in robust water body identification.

Keywords: Landsat 8; OLI sensor; J48 decision tree; water body identification

1. Introduction

Water is an essential component for the sustainability of life on earth. It is equally important to humans and other forms of life. Its presence and absence causes land use change. It balances ecosystem as well as maintains climate variation, carbon cycle etc. Hence, identification of such water bodies are essential and can be useful in various ways such as estimation of water availability [1], demarcation of flooded regions [2], wetland inventory [3] and so on.

With increased availability and improved quality of multi-temporal remote sensing data, identifying water bodies has been quite easier. In past decades, Landsat sensors have been used for land use classification using various unsupervised and supervise methods. With the introduction of new OLI sensor in Landsat 8 with improved qualities, the accuracy of classification has been much improved. Visual interpretation could easily provide highly accurate identification but consumes longer time in case of high resolution data. With the development of Geographic Information Science (GIS) and computer sciences, several contributions had been made for automatic feature identification. Some of the most well-known multiband water classification methods are: Normalized Difference Water Index (NDWI) [4], Modified NDWI (MNDWI) [5] and Automated Water Extraction Index (AWEI) [6].

With advancement in technology, the resolution of data are increasing, increasing in data size, and thus requiring more robust and faster methods of detection with high accuracy. For such problem, new data mining methods such as neural network, neurologic, decision tree, vector machines etc. are to be explored. These data mining techniques has been proposed, implemented and shown good identification capability in many other fields.

The main objective of this study is to apply J48 decision tree (JDT) model to identify water bodies using Landsat 8 OLI imagery. JDT is an open source java implementation of C4.5 decision tree [7]. The method is easy to understand and implement in GIS.

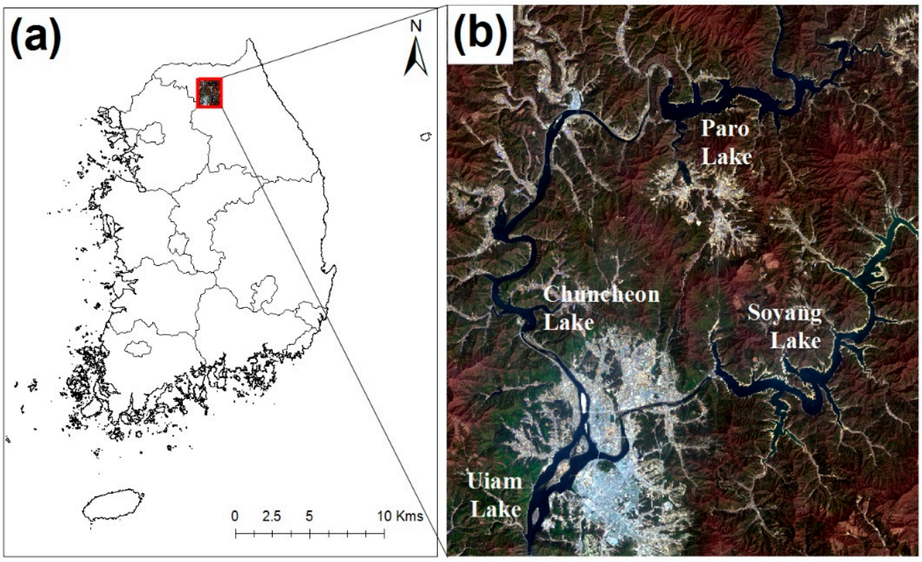

Figure 1. (**a**) Location of the test site in Korea. (**b**) Landsat natural colour composite imagery for the test site.

2. Experiment Section

2.1. Test Site

Water bodies in Han River basin, located in Gangwon province, Republic of Korea are selected as the test sites. The site area consists of four lakes: Uiam Lake (Chuncheon), Chuncheon Lake (Hwacheon), Paro Lake (Yanggu) and Soyang Lake (Inje). All the lake are formed by the respective dams in the river and are the main source of water supply to Seoul area. The area contains large water surface, thus it was selected for the testing of water body identification algorithm.

2.2. Data and Methods

The Level 1 Terrain Corrected (L1T) data acquired by Landsat 8 Operational Land Imager (OLI) sensor on 24 April, 2015 was used for the extraction of water bodies. It was collected from the Global visualization viewer (GLOVIS) of United States Geological survey (USGS) website. The obtained multiband image consisting coastal blue, blue, green, red, near infrared (NIR), shortwave infrared 1 (SWIR_1) and shortwave infrared 1 (SWIR_2), were converted to top-of-atmosphere (TOA) reflectance using the Landsat calibration tool in ENVI 4.8. The required information including the Data Acquisition Date and Sun Elevation was obtained from the Landsat MTL file. A 30 meters resolution scene of 1216 rows and 1054 columns were clipped for the study purpose. Each of the pixels with all band values was exported into Comma Separated Value (CSV) table for classification over Weka. Out of the whole scene, randomly 9800 pixels were sampled for training and validation of model development.

In Weka, JDT is used for development of model using 70% of sampled data. The model was cross validated using reaming 30% data for error estimation. The performance of the result was statically analyzed using *Kappa statistics* and *Area under Curve*. Upon the acceptable condition, the model would be applied to the whole scene and thus the binary non-water and water classification was done.

3. Results and Discussion

The JDT developed out of the training data is given in Figure 2. The size of the tree is 19 and consists of 10 leaves. From figure it can be clearly seen that most important classification role is played by NIR band followed by SWIR1, blue and red respectively whereas coastal blue and green were rejected from the decision tree.

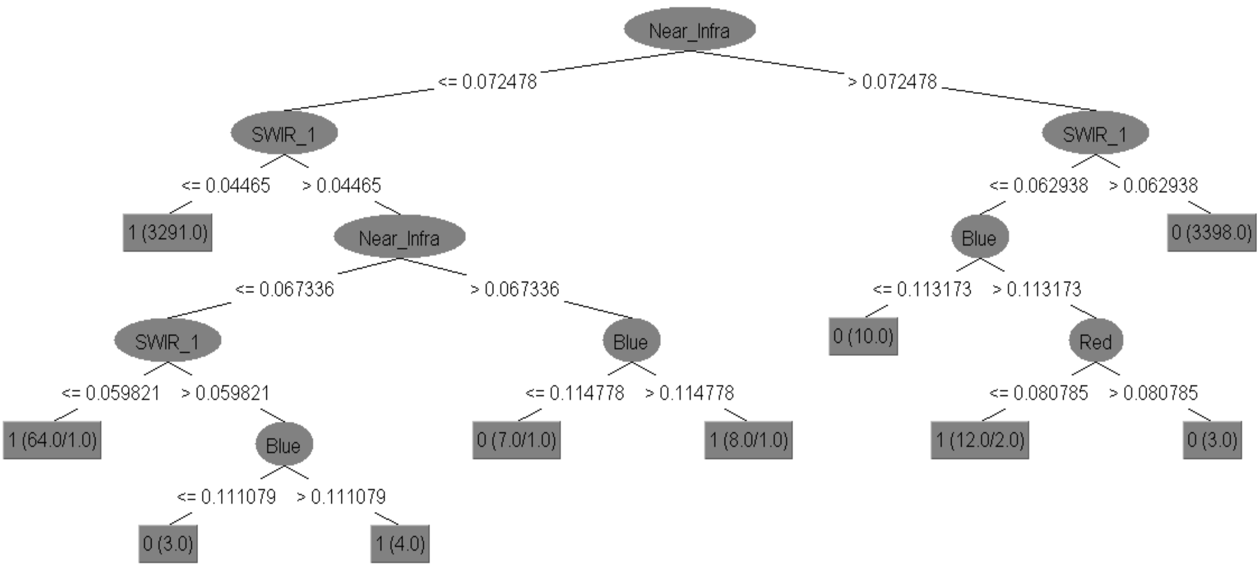

Figure 2. J48 decision tree model for water identification in the test site.

The sampled data were populated with binary non-water and water class using infrared band and high resolution imagery from Google Earth Pro. The JDT model classified instances 99.92 % correctly, had *Kappa Statistic* of 0.9985 and *Area under Curve (AUC)* was 1.0. Being the carefully chosen training data the accuracy was expected. Hence, the data unused for the training is very important to evaluate the model. The reaming 30% sampled instances were 99.4% correctly classified whereas showed *Kappa Statistics* of 0.9881 and *AUC* of 0.997. Both of the statistics indicate the high accuracy of classification of non-water and water pixels. The method is not only accurate but very fast also in processing. Hence, JDT is a very good method for water identification. Figure 3 shows the non-water and water body (blue) classification obtained from the JDT method.

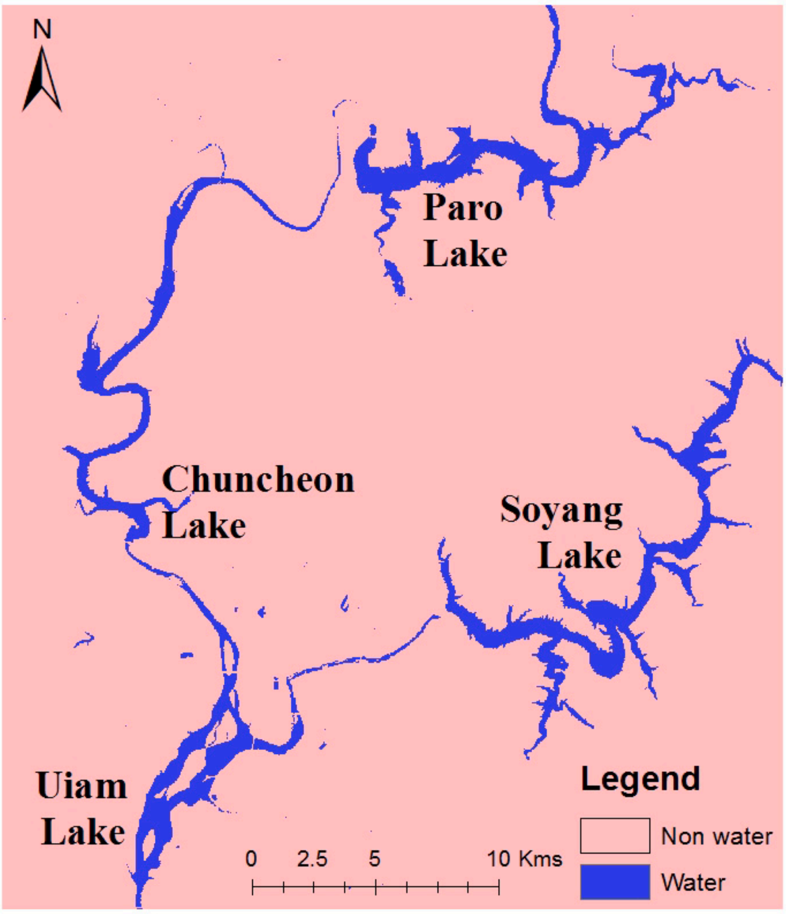

Figure 3. Water body identification result of the test site.

4. Conclusions

Water is an important part of the ecosystem. Identification of water is very important for various scientific estimation as well as social problem solving. In this current study, JDT was applied to identify the non-water and water bodies. Randomly sampled pixels were classified using NIR and high resolution imagery and then 70% were used for modeling whereas reaming for validation. In modeling and validation, the accuracy is very good in identification of non-water and water bodies. Hence, JDT could be a good tool for fast and accurate for water in cases like estimation of water availability, demarcation of flooded regions, wetland inventory and so on.

Author Contributions

Tri Dev Acharya performed and prepared this research article. In Tae Yang and Dong Ha Lee supervised and helped in manuscript writing and final submission.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Rover, J.; Ji, L.; Wylie, B.K.; Tieszen, L.L. Establishing Water Body Areal Extent Trends in Interior Alaska from Multi-Temporal Landsat Data. Remote Sensing Letters **2012**, *3*, 595-604.
2. Jain, S.; Singh, R.D.; Jain, M.K.; Lohani, A.K. Delineation of Flood-Prone Areas using Remote Sensing Techniques. Water Resour. Manage. **2005**, *19*, 333-347.
3. Rebelo, L.M.; Finlayson, C.M.; Nagabhatla, N. Remote Sensing and GIS for Wetland Inventory, Mapping and Change Analysis. J. Environ. Manage. **2009**, *90*, 2144-2153.
4. McFEETERS, S.K. The use of the Normalized Difference Water Index (NDWI) in the Delineation of Open Water Features. Int. J. Remote Sens. **1996**, *17*, 1425-1432.
5. Xu, H. Modification of Normalised Difference Water Index (NDWI) to Enhance Open Water Features in Remotely Sensed Imagery. Int. J. Remote Sens. **2006**, *27*, 3025-3033.
6. Feyisa, G.L.; Meilby, H.; Fensholt, R.; Proud, S.R. Automated Water Extraction Index: A New Technique for Surface Water Mapping using Landsat Imagery. Remote Sens. Environ. **2014**, *140*, 23-35.
7. Quinlan, J.R.; Cameron-Jones, R.M. FOIL: A Midterm Report. In Machine Learning: ECML-93; pp. 1-20.